



# Projet METAL : Fondements pour la conception d'un entrepôt de données d'apprentissage

Azim Roussanaly, Laura Infante Blanco

## ► To cite this version:

Azim Roussanaly, Laura Infante Blanco. Projet METAL : Fondements pour la conception d'un entrepôt de données d'apprentissage. [Rapport de recherche] Université de lorraine; CNRS. 2018. hal-01801899

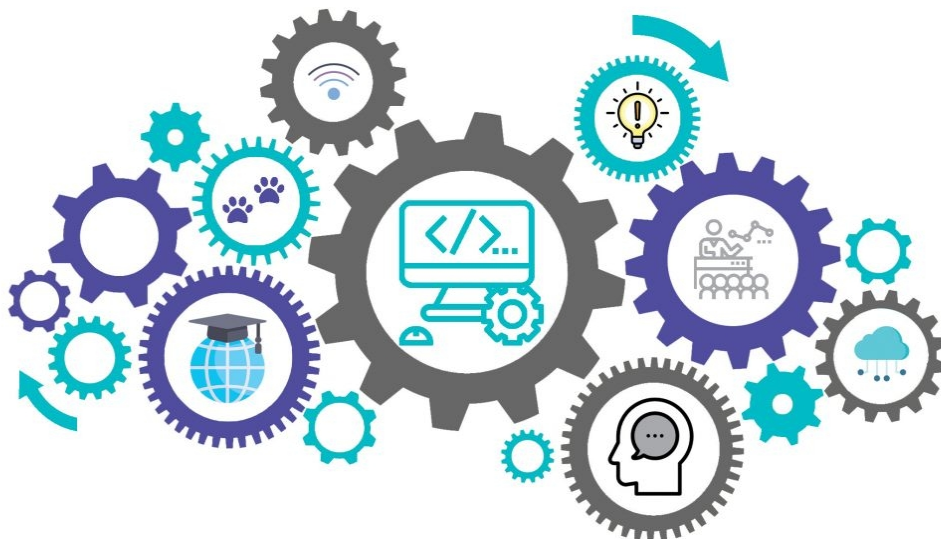
**HAL Id: hal-01801899**

**<https://inria.hal.science/hal-01801899>**

Submitted on 28 May 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Projet METAL

Modèles Et Traces au Service de l'Apprentissage des Langues

**Fondements pour la conception d'un entrepôt de données d'apprentissage**



Azim Roussanaly, Laura Infante Blanco

Équipe KIWI, LORIA, Université de Lorraine

25/04/2018 – v. 1,4

## Introduction

Ces dernières années, une production grandissante des données issues des environnements numériques éducatifs a justifié l'émergence et le croissant intérêt pour une nouvelle discipline de gestion et analyse des données massives en éducation appelée *Learning Analytics*<sup>1</sup>.

Les *Learning Analytics* ou **analyse des données d'apprentissage** tirent profit de la collecte de données de masse, pour la modélisation du contexte éducatif et son utilisation pratique par les différents acteurs de la communauté éducative : les enseignants bénéficieraient d'une meilleure compréhension des profils de leurs apprenants et, de ce fait, pourraient proposer des parcours alternatifs ou personnalisés (*adaptive learning*) ; les étudiants, quant à eux, seraient en mesure de maîtriser l'efficacité de leur apprentissage à l'aide des indicateurs de leur progression personnelle ; en plus grosse granularité, ces technologies pourraient venir en aide à l'évaluation du fonctionnement global des programmes pédagogiques.

Les *Learning Analytics* constituent un domaine de recherche relativement récent, aux contours encore imparfaitement définis<sup>2</sup>, mais héritant en particulier du domaine des *Data Analytics*. Ce domaine comprend des disciplines diverses comme les **statistiques, l'apprentissage automatique**<sup>3</sup> ou la **fouille de données**<sup>4</sup>, sur des données éducatives de grand taille. La **collecte de ces données**, non seulement fait partie de ce domaine de recherche, mais conditionne aussi de manière importante tous les autres traitements en aval et est estimée à 80% du volume du travail lorsqu'un système d'analyse de données est mis en place<sup>5</sup>.

La **collecte de données** de masse regroupe un ensemble d'activités comme **l'évaluation** des sources et la **génération, l'acquisition, le traitement** et le

---

<sup>1</sup>Siemens, George. "Learning Analytics: The Emergence of a Discipline" *American Behavioral Scientist*, vol. 57, no. 10, 2013, pp. 1380–1400., doi:10.1177/0002764213498851.

<sup>2</sup>Ferguson, Rebecca. "Learning Analytics: Drivers, Developments and Challenges." *International Journal of Technology Enhanced Learning*, vol. 4, no. 5/6, 2012, p. 304., doi:10.1504/ijtel.2012.051816.

<sup>3</sup> Plus connu sur le nom de **Machine Learning**

<sup>4</sup>Siemens, George, and Ryan S. J. D. Baker. "Learning Analytics and Educational Data Mining." *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 2012, doi:10.1145/2330601.2330661.

<sup>5</sup> [http://visit.crowdflower.com/rs/416-ZBE-142/images/CrowdFlower\\_DataScienceReport\\_2016.pdf](http://visit.crowdflower.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf)

**stockage** des données sans oublier la sécurité des données et la protection des données sensibles<sup>6</sup>.

Dans le cadre du projet **METAL**<sup>7</sup> (Modèles Et Traces Au service de l'apprentissage des Langues), qui, rappelons-le, ambitionne de concevoir et d'expérimenter des services de *Learning Analytics* au bénéfice des établissements secondaires, nous avons évidemment été confrontés à cette problématique avec les spécificités qui caractérisent le système éducatif en France, voire dans une académie.

Le travail en cours nous a amenés à définir un **modèle des données**, sur lequel s'appuie la conception d'un entrepôt de données d'apprentissage (*LRS*), destiné à collecter, à stocker et à mettre à la disposition des **applications du projet METAL** l'ensemble des données. Dans ce document, nous proposons de présenter synthétiquement notre approche, dans une perspective de généralisation de notre méthodologie à travers l'exemple de l'Académie de Nancy-Metz. Pour ce faire, nous allons aborder le sujet à travers quatre questions fondamentales auxquelles nous avons été confrontés lors de notre étude :

1. Quelles données ?
2. Qui les détient ?
3. Qui les fournit ?
4. Où les stocker ?

## Quelles données ?

Le recensement des données disponibles ou susceptibles d'être disponibles sous forme numérique s'effectue sur la base de la compilation de différentes observations qui s'appuient essentiellement sur des entrevues avec différents acteurs, sur l'observation des usages des applications en situation réelle et enfin sur un travail de documentation sur le système éducatif, plus spécifiquement, sur ce qui concerne le collège, cible de notre expérimentation.

---

<sup>6</sup>H.V. Jagadish, Johannes Gehrke, Alexandros Labrinidis, Yannis Papakonstantinou, Jignesh M. Patel, Raghu Ramakrishnan, Cyrus Shahabi, *Big data and its technical challenges*, *Commun. ACM* 57 (7) (July 2014) 86–94, <http://dx.doi.org/10.1145/2611567>

<sup>7</sup> <http://metal.loria.fr>

En effet, nous avons eu accès à des démonstrations des usages des services disponibles à travers l'espace numérique de travail (ENT) de la part d'enseignants et de parents d'élèves. En l'occurrence, il s'agit principalement de l'ENT PLACE<sup>8</sup>, dont le déploiement a été généralisé à l'échelle de l'Académie de Nancy-Metz en 2013. Par ailleurs, nous avons également organisé une série d'échanges avec des représentants de l'éditeur de l'ENT (dans notre cas ITOP Éducation, partenaire du projet METAL) et les fournisseurs des infrastructures (dans notre cas les Directions des Systèmes d'Information (DSI), des Conseils Départementaux et du Rectorat de l'Académie).

A l'issue de ce recensement, nous distinguons **4 catégories de données** caractérisées selon leur nature, en l'occurrence, 1) les données relatives à la **vie scolaire**, 2) les données d'**utilisation des services**, 3) les données d'**usages des ressources** pédagogiques et enfin 4) les données de **nature démographique**, que nous allons maintenant détailler.

## Les données de la vie scolaire

Ces données comprennent des informations dont la fréquence de production est très variable, dès celles générées tous les jours comme les absences et incidences jusqu'à celles qui ne sont produites qu'une seule fois tous les trois ou quatre ans comme les bilans de cycle. Nous les énumérons à continuation selon sa périodicité de génération.

Il y a tout d'abord les informations liées à l'organisation pédagogique, qui constituent nos **référentiels** et qui ont la particularité d'être statiques dans le temps, c'est-à-dire, des informations qui ne sont modifiées que par périodes de plusieurs années, lors des changements de lois ou des décrets de l'éducation. On y trouve notamment les informations concernant les **cycles pédagogiques**, les **enseignements annuels** et parmi ceux-ci, les **enseignements facultatifs**, les enseignements pratiques interdisciplinaires (**EPI**) et les accompagnements personnalisés (**AP**)<sup>9</sup>.

La vie scolaire comprend également d'autres informations dont la périodicité est de quelques années, comme les **bilans de cycle**, mais elles ne font pas partie des référentiels parce qu'elles font référence à un élève en particulier et non pas à tout le système éducatif en soi.

<sup>8</sup> <http://ent-place.fr/le-projet/>

<sup>9</sup> <http://www.education.gouv.fr/cid214/le-college-enseignements-organisation-et-fonctionnement.html>

Ensuite nous avons des documents dont la périodicité est annuelle ou trimestrielle, comme les **bilans périodiques**, où nous retrouvons les **moyennes** des élèves pour le semestre ou l'année, ainsi que d'autres informations complémentaires en forme **d'observations** des enseignants. Pour finir, nous avons identifié des données avec une haute fréquence de production, comme par exemple les **absences**, les **incidents**, les **devoirs**, les **notes** des détaillées, etc.

## Les données d'utilisation de services

A ces informations, essentiellement fournies par l'administration et les enseignants, viennent s'ajouter les informations engendrées par les élèves ou par leurs parents lors de leur utilisation des services en ligne. Celles-ci permettent, par exemple, de déterminer si un collégien a consulté son **emploi du temps** de la semaine ou si un parent d'élève a visualisé le **cahier de texte** de son enfant.

## Les données d'usage des ressources pédagogiques

Nous entendons par **ressource pédagogique** tout dispositif permettant à un enseignant de diffuser des documents numériques (exemple : encyclopédie, ouvrages, documents multimédia...), d'interagir avec un élève ou un groupe d'élèves (exemple : QCM, ressources interactives...) ou de mettre à la disposition de ces derniers un espace collaboratif en ligne (exemple : rendu de travaux, évaluation...).

Ces contenus peuvent être proposés aux étudiants sous la forme d'unités pédagogiques dans le cadre d'un élément d'un programme ou sous la forme de parcours d'activités personnalisés où une unité s'enchaînera à une autre éventuellement sous certaines conditions.

Dans cette catégorie, nous recueillons d'une part, la **description des ressources pédagogiques** et, d'autre part, les **traces d'usage** engendrées par l'utilisation individuelle de ces ressources.

La description des ressources comprend notamment des éléments généraux comme les titres, les résumés, les relations avec les enseignements concernés, mais aussi des caractéristiques d'usage comme par exemple les notes minimales et maximales pour une activité notée ou la durée pour une vidéo.



Les traces d'usage des apprenants sur ces ressources décrivent l'interaction des élèves avec ces éléments. Par exemple, dans le cas d'une ressource vidéo, nous enregistrons notamment la date d'accès et la durée de visionnage.

## Les données de nature démographique

Ces données concernent non seulement les élèves, mais aussi leur famille. Ces informations ont une périodicité de collecte longue, voire permanente, comme dans le cas de la **nationalité** ou le **lieu de naissance**. Autres données susceptibles d'appartenir à cette catégorie sont l'**adresse**, le **genre**, la **date de naissance**, l'**état civil des parents**, la **date de naissance des parents**, leur **catégorie socioprofessionnelle**, etc.

Notons que, de part le caractère personnel de ces données, leur collecte, leur stockage et leur traitement entre dans le cadre du règlement général sur la protection des données personnelles qui nécessite le consentement éclairé des individus concernés ou de leurs représentants légaux.

La démarche que nous venons de présenter est résolument **ascendante**, c'est-à-dire qu'elle s'appuie sur "**l'observation de terrain**" pour identifier et organiser les données à recueillir. Comme nous le verrons plus bas, cette démarche est combinée avec une approche descendante, basée sur l'examen des normes dans le domaine, afin d'établir un modèle conceptuel des données à recueillir.

## Qui détient les données ?

Nous définissons les acteurs qui détiennent les données comme étant les organisations **responsables** juridiquement des données que nous cherchons à collecter. Par conséquent, c'est auprès de celles-ci que nous devons solliciter le droit de disposer des données en conformité avec la réglementation.

Pour le projet METAL, cela concerne les **chefs d'établissements** impliqués dans l'expérimentation, le **Rectorat** et le **Conseil Départemental** de Moselle (car tous les établissements expérimentateurs se trouvent dans ce département).

Les acteurs, comme la société ITOP, qui déploient les services en ligne, potentiellement source de données, sont considérés comme étant des **prestataires**.

Et enfin, les élèves, leurs parents et les enseignants appartiennent à la catégorie des **producteurs de données**.

Le processus de collecte, de stockage et de traitement des données, impose une identification claire des rôles liés à différents niveaux d'obligation réglementaires. D'autres caractéristiques, que nous listons ci-dessous, interviennent également dans le processus de mise en place de mesures de protection des données :

- Le projet METAL est considéré comme un projet de recherche, et, de ce fait, entre dans un cadre réglementaire dérogatoire spécifique.
- De part la nature même des recherches menées, intégrant la recherche de combinaisons optimales des données pour améliorer les performances des algorithmes, le principe de minimisation des données collectées pour les besoins des traitements reste complexe à définir *a priori*. Cependant, le projet se plaçant sur le champ du *Learning Analytics* (Analyse de l'apprentissage) et non pas du *Educational Analytics* (Analyse des données d'éducation), seules les données à caractère pédagogique en relation avec le processus d'apprentissage, produites par les élèves, leurs parents et les enseignants sont collectées. Seront exclues, en conséquence, les données produites par d'autres acteurs comme par exemple le personnel administratif.
- Les données personnelles sont utilisées pour des nouveaux traitements (autres que ceux pour lesquels ils ont été prévus initialement). Cela implique la mise en place de modalités du recueil du consentement des acteurs concernés.

Dans le cadre du projet METAL, c'est le Rectorat de l'Académie de Nancy-Metz, par l'intermédiaire de la Délégation Académique au Numérique pour l'Éducation (DANE) qui coordonne la problématique du droit et de la protection des données personnelles.



## Qui fournit les données ?

D'une manière générale, les sources de données sont étroitement liées aux différentes applications et services déployés à l'intention des élèves, des parents et des enseignants.

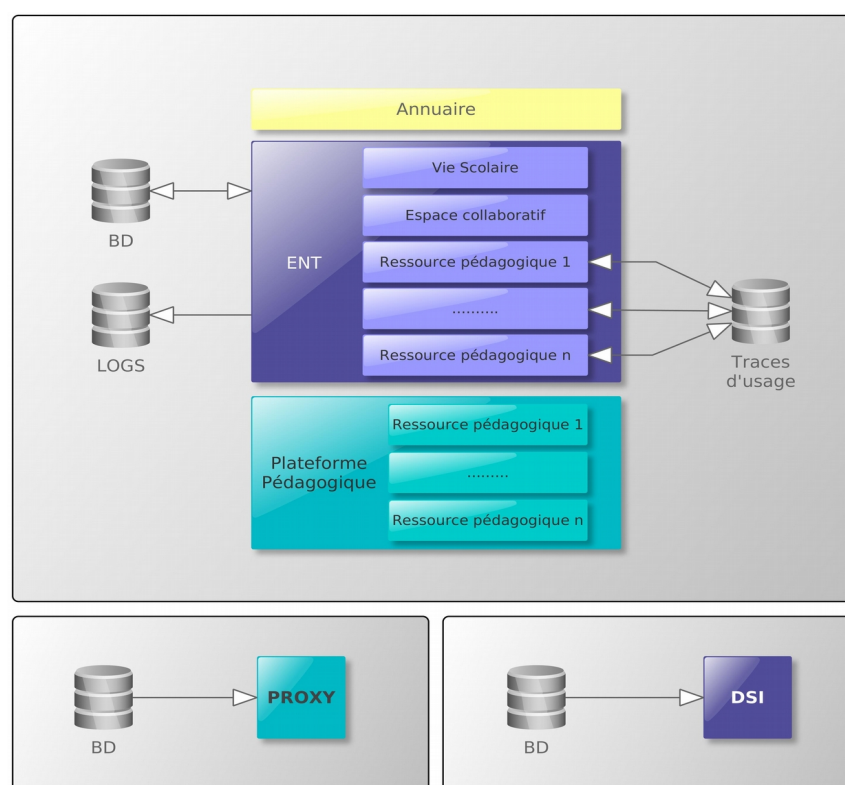


Figure1: potentielles sources de données pour METAL

La figure 1 fournit schématiquement l'ensemble des sources disponibles susceptibles de procurer les données. Nous distinguons plusieurs sources potentielles :

- **L'ENT** et les composantes intégrées sous la forme d'extensions articulées autour d'un **annuaire** commun. Dans la version actuelle (c'est-à-dire l'ENT PLACE) un module de **vie scolaire** et un **espace collaboratif** sont généralement déployés de manière standard dans les établissements

secondaires. Pour des raisons d'obligation de traçabilité, l'ENT génère de fichiers de journalisation des activités (**LOG**) qui représentent une source de données importante pour le projet METAL.

- De manière complémentaire, on trouve, selon les établissements, des **ressources pédagogiques** numériques, produites par différents éditeurs, implantées sous la forme d'extensions standardisées, s'appuyant sur l'annuaire commun de l'ENT et interagissant éventuellement avec les autres services de l'ENT. Certaines de ces ressources embarquent, de manière autonome, un module permettant de collecter des **traces d'usage** liées aux utilisations spécifiques de la ressource. Ces traces constituent également une source de données pour le projet METAL.
- D'autres **plateformes pédagogiques**, généralement des LMS (*Learning Management System*) proposés par les Banques de Ressources Numériques pour l'École (BRNE)<sup>10</sup>, offrent également des accès à des **ressources pédagogiques**. A ce stade du projet, nous n'avons pas accès aux traces d'usages de ces ressources, en raison, d'une part de leur grande diversité et, d'autre part, de la multiplication des annuaires, rendant très complexe l'identification des apprenants concernés. Nous mentionnons ici leur existence, en espérant, à terme, envisager également la collecte des données provenant de ces plateformes grâce notamment au projet national d'unification des annuaires au niveau des instances nationales.
- Le **Proxy** : comme décrit plus haut, pour des motivations de gestion et de sécurité, le Conseil de département, responsable des infrastructures, met en place un Proxy, dont la fonctionnalité principale est de canaliser les échanges entre le réseau local et les réseaux distants, minimisant ainsi les points de surveillance du trafic numérique. Un proxy a la faculté de journaliser les différents accès des usagers aux différents services. Ces informations constituent une source de données supplémentaire pour le projet METAL. Nous n'utilisons pas ici la totalité de ces informations, mais nous limitons les besoins aux activités des enseignants et des élèves portant sur des ressources préalablement répertoriées.
- La **DSI** : cette entité représente globalement le système d'information du rectorat, et plus précisément le volet qui traite des utilisateurs et des résultats académiques. Par conséquent, le système d'information constitue une source importante de données permettant des informations

<sup>10</sup> <http://ecolenumerique.education.gouv.fr/brne/>

démographiques sur les élèves et leurs parents ainsi que sur l'historique des résultats de chaque élève.

## Où stocker les données ?

Concrètement, le processus de collecte de données consiste à alimenter un entrepôt de données destiné à stocker de manière centralisée l'ensemble des données sur lesquelles s'appuient les traitements de *Learning Analytics*.

### Les LRS

Un *LRS* (*Learning Record Store*) ou *entrepôt de données d'apprentissage* est un système de stockage centralisé qui permet d'enregistrer des usages de ressources issues de multiples systèmes d'apprentissage en ligne avec l'objectif de proposer des technologies du *Learning Analytics*. Ces usages de ressources, appelés *statements* dans la communauté de l'apprentissage numérique, sont exprimés en forme de triplets <acteur, verbe, objet>. Pour illustrer ceci de façon simplifiée, le *statement* qui serait enregistré quand Pierre (un étudiant de cinquième) regarde une vidéo sur les équations de premier degré, serait le triplet : <Pierre, a regardé, systèmes d'équations de premier degré> (voir figure 2).

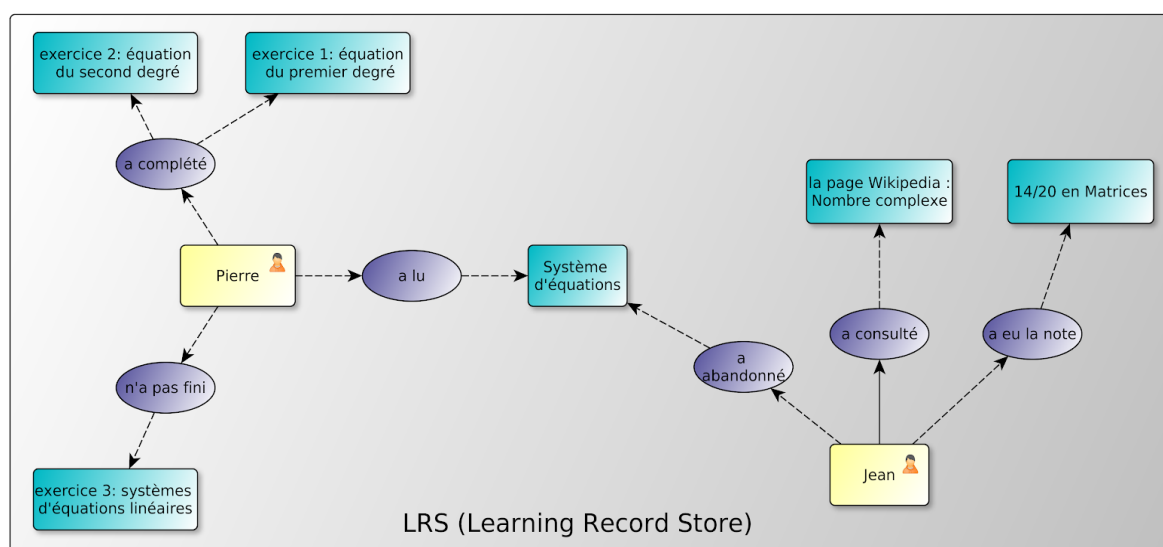


Figure 2: exemple de xAPI statements dans un LRS

Il existe plusieurs standards internationaux de spécification de ces systèmes de stockage de données d'apprentissage, les plus répandues étant *experience API* (xAPI)<sup>11</sup> de l'*Advanced Distributed Learning Initiative* (ADL)<sup>12</sup>, un programme américain pour la recherche et développement de l'apprentissage numérique) et *Caliper Analytics*<sup>13</sup>, un standard de *IMS Global*<sup>14</sup>, une organisation internationale à but non-lucrative pour les technologies de l'apprentissage.

## Les traces d'usage : le standard xAPI

Les environnements numériques de travail les plus utilisés dans le monde, comme *Moodle*<sup>15</sup>, *Sakai*<sup>16</sup>, *eTutor*<sup>17</sup> ou *OLAT*<sup>18</sup>, implémentent des *connecteurs* xAPI, qu'ils soient intégrés directement dans le code des versions plus récentes, ou bien sous forme d'extension à installer sur les plateformes. Un *connecteur* ou *producteur* xAPI est un logiciel qui permet de repérer les actions des apprenants, les transformer en *triplets* ou *statements* et les envoyer au LRS. Le LRS de son côté implémente un *consommateur* xAPI, qui à son tour, recevra les contenus des triplets provenant des différents systèmes d'apprentissage en ligne pour y être stockés. Ces données serviront ensuite à la construction de modèles d'utilisateurs visant à améliorer l'expérience d'apprentissage des étudiants et la compréhension du profil d'apprentissage des élèves de la part de leurs enseignants.

Nous avons choisi d'utiliser ce standard pour notre LRS en raison, d'une part, de la disponibilité de producteurs xAPI dans le système de ressources en ligne de l'ENT actuel (*NetEduc Cloud*), et, d'autre part, du futur prometteur de ce standard adopté par de plus en plus d'éditeurs de ressources pédagogiques numériques en France et dans le monde.

Mais METAL LRS veut aller un peu plus loin et nous ne nous contentons pas du stockage des traces d'usage des ressources pédagogiques numériques. En

---

<sup>11</sup> <https://xapi.com/overview/>

<sup>12</sup> <https://adlnet.gov/>

<sup>13</sup> <https://www.imsglobal.org/activity/caliper/>

<sup>14</sup> <https://www.imsglobal.org/>

<sup>15</sup> <https://www.moodle.org/>

<sup>16</sup> <https://www.sakaiproject.org/>

<sup>17</sup> <http://www.e-tutor.com/>

<sup>18</sup> <https://olat.org/>

effet, d'autres activités plus traditionnelles, engendrent des informations indirectes liées au comportement pédagogique d'un élève comme par exemple les notes, les absences ou les incidents. Il apparaît important que ces informations soient collectées au même titre que les traces d'usage d'une ressource. Nous proposons donc une extension du vocabulaire pour les verbes *xAPI* qui permettrait de codifier des actions comme s'absenter d'un cours physique, de rendre un devoir non numérique, ou de l'évaluation d'un partiel sur papier.

Ensuite, nous considérons que les référentiels dont on a parlé dans les sections antérieures, nous permettent de mettre en contexte les traces d'utilisation de services et d'usage des ressources numériques. Ces données, n'ont pas de place dans un *LRS* construit avec le standard *xAPI*. Nous avons donc décidé de combiner ce standard, pour le stockage de traces d'usage, avec le standard *OneRoster* de *IMS Global*, qui nous permet de stocker les référentiels, et des données de la vie scolaire comme les inscriptions, la démographie ou l'emploi du temps des élèves.

## La vie scolaire : *OneRoster*

Ce standard international de la fondation *IMS Global*<sup>19</sup> vise au partage et la transmission des unités d'information appelées *rosters* entre différents systèmes informatiques, par exemple, entre *l'ENT* et le *LRS*. Un *roster* représente une liste d'informations, normalement reliées à un individu, e.g. une liste de cours, d'inscriptions, d'enseignants, etc.

La spécification *OneRoster* définit en détail les principaux éléments de la **vie scolaire**. On y retrouve ainsi de manière détaillée les concepts éducatifs suivants : l'établissement, la classe, le cours, l'inscription, la période d'évaluation, l'élève, l'enseignant, la démographie, l'emploi du temps, la ressources éducative et le résultat académique.

Ce modèle est présenté en **deux modalités** pour l'échange : la première permet le transfert des données en forme de fichier **CSV** (*Comma Separated Value*), et la deuxième décrit les services d'une éventuelle **API**<sup>20</sup> pour l'envoi des données sur le protocole *HTTP*<sup>21</sup>.

---

<sup>19</sup> <https://www.imsglobal.org/>

<sup>20</sup> Acronyme pour *Application Programming Interface* (interface de programmation applicative) désigne l'ensemble des services offerts par un logiciel à d'autres logiciels

<sup>21</sup> Abréviation de *Hypertext Transfer Protocol*, protocole de communication client-serveur utilisé dans le Web

L'adoption, dans un premier temps du modèle de transfert de fichiers CSV, présente l'avantage de parvenir à une implantation plus rapide, mais cela comporte l'inconvénient de nécessiter une intervention humaine pour la génération de ces fichiers, et aussi pour leur importation. Dans un deuxième temps, idéalement, il serait plus efficace de pourvoir les services de *sensors*. Un *sensor* en terminologie *OneRoster*, est un capteur, associé par exemple à un *ENT* ou à un autre système de gestion ou d'apprentissage en ligne, capable d'extraire les informations nécessaires pour conformer les éléments des concepts du modèle et de les envoyer automatiquement à travers le protocole *HTTP* au *LRS*. Cette modalité permettrait d'automatiser la tâche et d'éviter ainsi des tâches humaines monotones et répétitives.

## METAL LRS

Notre *LRS* se base principalement dans l'adoption des standards *xAPI* et *OneRoster*. Ce modèle hybride a pour but l'utilisation des meilleurs caractéristiques de chaque standard, tout en les adaptant au modèle de l'enseignement secondaire français.

*xAPI* apporte un standard pour le recueil de traces d'apprentissage, et nous l'avons enrichi de nouvelles actions pour acquérir également d'autres actions non prévues actuellement dans le standard.

*OneRoster* est un bon départ pour la normalisation des données de vie scolaire, démographiques et référentiels. Mais nous avons dû modifier le modèle en raison notamment des divergences avec le modèle américain concernant la démographie des étudiants et la description de l'emploi du temps. Nous avons également remplacé toute la partie destinée à collecter les traces d'apprentissage (que nous avons jugée trop rudimentaire) par *xAPI* qui est un standard bien plus puissant et répandu dans les *LMS* ou *ENT's*.

Un modèle conceptuel de données final synthétique pour le *LRS* de METAL est présenté dans la figure 3 :



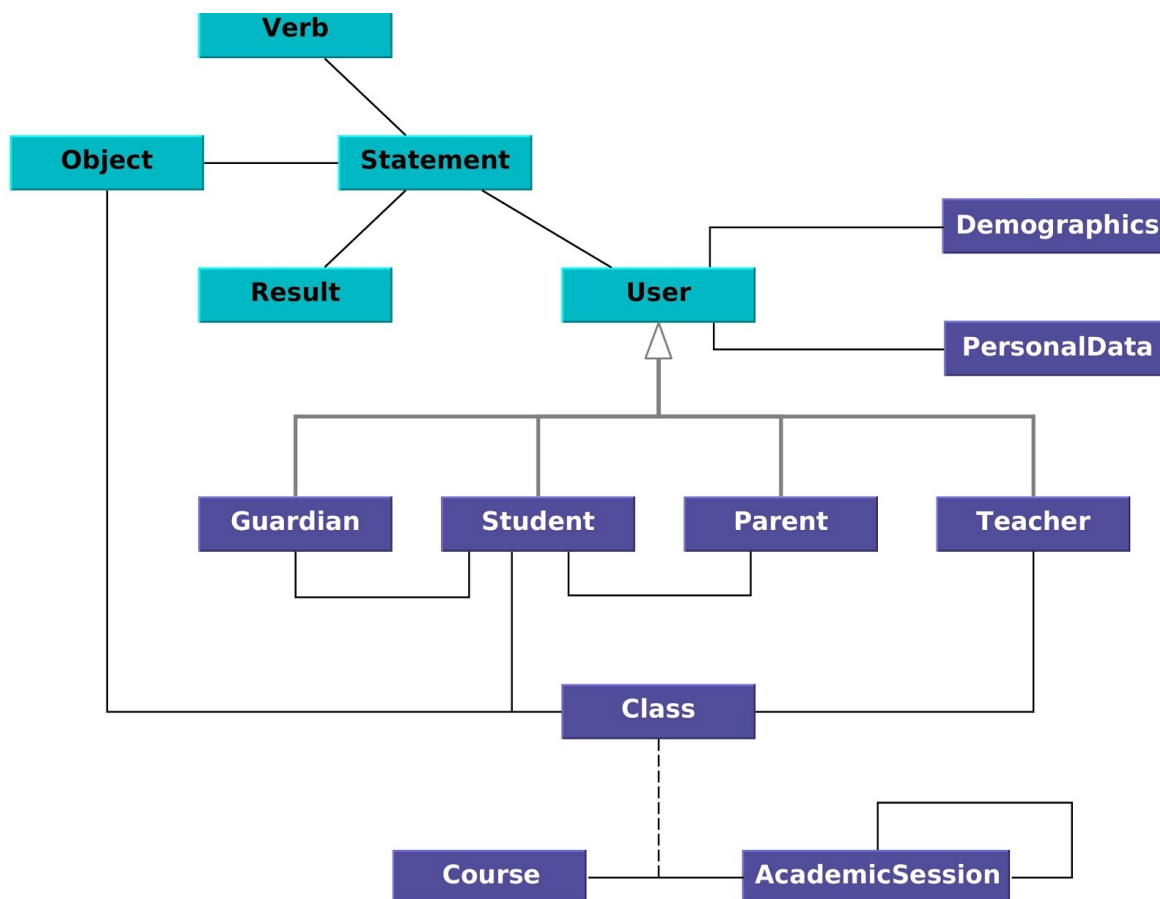


Figure 3: Modèle conceptuel de données de METAL LRS

## Conclusion

À travers cette étude de cas et les quatre questions traitées dans ce document, nous mettons en évidence l'existence de plusieurs dimensions dans une démarche méthodologique pour un projet de conception d'un entrepôt de données d'apprentissage (LRS).

La première dimension concerne la structure même des informations contenues dans un LRS. Dans notre exemple, nous avons abordé le problème selon un double point de vue. Le premier se veut ascendant, c'est-à-dire, en s'appuyant sur des informations disponibles, identifiées à partir d'une observation de terrain (*in vivo*). Le second se veut descendant, c'est-à-dire, à partir des

standards existants dans le domaine de l'apprentissage. C'est de la combinaison de ces deux points de vue que nous définissons le modèle de données du *METAL LRS*, qui se distingue de *OneRoster* essentiellement, d'une part, par la prise en compte du contexte de l'enseignement secondaire et de la réglementation nationale, et, d'autre part, par une volonté d'intégration plus précise du standard xAPI, en vue de faciliter les spécifications fonctionnelles des accès à l'entrepôt pour les futures applications de *Learning Analytics*.

La deuxième dimension est de nature réglementaire. La réglementation sur la protection des données personnelles définit un environnement dans lequel il est nécessaire de situer les traitements au sein du projet METAL. Cela conduit à une caractérisation du projet : (1) c'est un projet de recherche, (2) il concerne des données personnelles impliquant des personnes mineures, (3) il collecte des données pour un usage spécifique d'applications de *Learning Analytics* et (4) ces données proviennent de sources multiples impliquant des organismes responsables et des prestataires définis. Ces caractéristiques constituent un cadre permettant de définir les obligations réglementaires pour la collecte et le traitement des données.

La troisième dimension est d'ordre organisationnelle. En effet, à ce stade nous sommes confrontés à plusieurs organismes susceptibles d'alimenter l'entrepôt. Il est probable qu'à moyen terme, on en identifie de nouveaux. L'organisation mise en place au sein du projet METAL, basée sur la coordination centralisée des relations avec les fournisseurs de données prise en charge par la DANE, répond à la problématique de la sélection cohérente de multiples sources de données et au respect du cadre éthique et réglementaire.

À ce stade, nous avons posé les fondements pour la conception de METAL LRS, l'entrepôt de données du projet METAL. Les étapes suivantes à court terme concernent :

- Le développement de l'entrepôt en intégrant les problématiques technologiques inhérentes au domaine abordé, liés à la complexité, la massivité, la sécurisation et à l'accès aux données. Ce travail est en cours.
- La définition des connecteurs destinés à alimenter l'entrepôt et à être déployés dans les applications tierces. Selon la temporalité du recueil des données, nous distinguons deux types de connecteurs : le premier, que nous appelons le capteur, qui communique en temps réel les données d'usage des applications et, le second, qui procède par extraction et importation de données, à des fréquences variables.

- L'amorçage de l'alimentation de l'entrepôt concerne la collecte et la mise en forme des anciennes données déjà présentes afin de fournir un minimum de données le plus rapidement possible pour les expérimentations de faisabilité des applications de *Learning Analytics* prévues dans le calendrier du projet METAL.